Chapter 3

The mouse as a model for human biology: a resource guide for complex traits analysis

Karen L. Svenson, Raymond F. Robledo, Carol J. Bult, Gary A. Churchill, Beverly J. Paigen and Luanne L. Peters

Nature Reviews Genetic 8(1): 58-69, 2007

The mouse as a model for human biology: a resource guide for complex trait analysis

Luanne L. Peters, Raymond F. Robledo, Carol J. Bult, Gary A. Churchill, Beverly J. Paigen and Karen L. Svenson

Abstract | The mouse has been a powerful force in elucidating the genetic basis of human physiology and pathophysiology. From its beginnings as the model organism for cancer research and transplantation biology to the present, when dissection of the genetic basis of complex disease is at the forefront of genomics research, an enormous and remarkable mouse resource infrastructure has accumulated. This review summarizes those resources and provides practical guidelines for their use, particularly in the analysis of quantitative traits.

Congenic Strain
A strain that is produced by repeated backcrossing (ten generations) to an inbred strain, with selection for heterozygosity at each generation for a specific locus in the donor strain.

In 1937, Life Magazine[1] noted the "similar physiological structure" of mice and humans, with the comparison: "Its life cycle is man's life cycle in miniature." We can now appreciate how visionary that statement really was, as we now know that mice and humans share ~99% of their genes[2]. This knowledge would probably have surprised even the most ardent 'mouser' of 1937. It follows that humans and mice also share common inherited diseases, both Mendelian and polygenic. These include diabetes, atherosclerosis, heart disease, cancer, glaucoma, anaemia, hypertension, obesity, osteoporosis, bleeding disorders, asthma and neurological disorders. Of course, other mammals also share these diseases. Still, the mouse is the premier model for the genetic basis of human disease because it is small, can be maintained cost-effectively, and has a short gestation period (~19–20 days). However, more significant is the wealth of information resources and experimental approaches for mouse genetics that have been systematically built up over the past 100 years. These resources and tools provide the means for understanding the genetic, molecular and cellular basis of human disease and normal biological processes.

After the establishment of the first inbred mouse strain (DBA) in the early 1900s, inbred strains were most prominent in cancer research and transplantation biology, for which genetically uniform stocks were required. Subsequently, the analysis of single-gene defects arising by spontaneous mutations in inbred strains moved to the forefront of mouse genetics research (see REFS 3,4 for excellent historical reviews of the mouse in biomedical research). Today, a third major historical stage in mouse genetics research is in full gear — complex, or quantitative, trait analysis. Unlike discrete traits (for example, eye colour, or the presence or absence of disease), quantitative traits (blood pressure, high density lipoprotein (HDL) levels, weight and so on) vary continuously over a range of distribution in a population and are influenced by multiple genes, as well as gene–gene and gene–environment interactions. As most common human diseases are complex, the identification of genes that influence such traits — QTL analysis — is the clear challenge for genetics in the future. It is noteworthy that the first two historical stages in mouse genetics research relied on reducing variation by inbreeding mice to homozygosity and producing inbred strains, whereas QTL analysis requires us to do the opposite — create and harness genotypic (and therefore phenotypic) variation to the fullest possible extent.

This review provides an overview of the available resources and tools in mouse genetics (BOX 1). We discuss: the impact of genetically pure inbred strains and their derivatives (for example, congenic strains and recombinant inbred strains) in the elucidation of basic biological processes and mechanisms of disease; genotype- and phenotype-driven strategies in gene identification and the elucidation of gene function; the ever-expanding catalogue of sequence and phenotypic information; and how the convergence of genome and phenome is empowering the dissection of complex traits. We provide appropriate database information and examples of how both animal and bioinformatic resources are used to identify the genes that underlie QTLs. We end with a description of the Collaborative Cross, an ambitious new initiative that will be an invaluable resource for complex genetics.

The Jackson Laboratory, 600 Main Street, Bar Harbor, Maine 04609, USA. Correspondence to L.L.P. e-mail: luanne.peters@jax.org
doi:10.1038/nrg2025

Box 1 | **Online resources for the mouse**

Mouse strain information and resources (inbred and mutant)

International Mouse Strain Resources	http://www.informatics.jax.org/imsr/index.jsp
JAX Mice	http://jaxmice.jax.org/index.html
Federation of International Mouse Resources	http://www.fimre.org
Mouse Mutant Resource	http://www.jax.org/mmr/index.html
Mouse Mutant Regional Resource Centers	http://www.mmrrc.org
Riken Bioresource Center	http://www.brc.riken.jp/lab/animal/en
The European Mouse Archive	http://www.emmanet.org
Mouse Models of Human Cancer Consortium	http:// http://mouse.ncifcrf.gov
Canadian Mouse Mutant Repository	http://www.cmmr.ca/index.html

Knockout and transgenic mice

International Gene Trap Consortium*	http://www.genetrap.org
Mouse Genome Infomatics Deltagen/Lexicon	http://www.informatics.jax.org
Induced Mutant Resource, Jackson Laboratory	http://www.jax.org/imr/index.html
Samuel Lunenfeld Research Institute‡	http://www.mshri.on.ca/nagy
Mouse Mutant Regional Resource Centers	http://www.mmrrc.org
Micer	http://www.sanger.ac.uk/PostGenomics/mousegenomics

Sequence/phenotype databases

Ensembl	http://www.ensembl.org/Mus_musculus/index.html
Map Viewer at NCBI	http://www.ncbi.nlm.nih.gov/mapview
Genome Browser, UCSC	http://genome.ucsc.edu/cgi-bin/hgGateway
Mouse Genome Informatics Database	http://www.informatics.jax.org
Vertebrate Genome Annotation	http://vega.sanger.ac.uk/index.html
Panther	http://pantherdb.org
Mouse Phenome Database	http://www.jax.org/phenome
Eumorphia	http://www.eumorphia.org
Mouse Tumor Biology Database	http://tumor.informatics.jax.org
German Mouse Clinic	http://www.gsf.de/ieg/gmc

Pathways analysis

Ingenuity	http://www.ingenuity.com
GenMAPP	http://www.genmapp.org
KEGG Pathway Database	http://www.genome.jp/kegg/pathway.html

SNP databases

Roche	http://mousesnp.roche.com
GNF	http://snp.gnf.org
NCBI	http://www.ncbi.nlm.nih.gov/SNP
Mouse Phenome Database	http://aretha.jax.org/pub-cgi/phenome/mpdcgi?rtn=snps/door
Mouse Genome Informatics Database	http://www.informatics.jax.org/menus/strain_menu.shtml
Perlegen	http://mouse.perlegen.com/mouse
Wellcome Trust Centre for Human Genetics	http://www.well.ox.ac.uk/mouse/INBREDS
Broad Institute	http://www.broad.mit.edu/personal/claire/MouseHapMap/Inbred.htm

Expression databases

GNF SymAtlas	http://symatlas.gnf.org
Institute for Genomic Research	http://pga.tigr.org
Gene Expression Omnibus	http://www.ncbi.nlm.nih.gov/geo
The Jackson Laboratory	http://www.informatics.jax.org/menus/expression_menu.shtml
Brain Atlas	http://www.brainatlas.org
GenSat	http://www.gensat.org/index.html
EMAGE	http://genex.hgu.mrc.ac.uk/Emage/database/emageIntro.html

Comparative genomics

VISTA	http://genome.lbl.gov/vista/index.shtml
Mouse Genome Informatics Database	http://www.informatics.jax.org/menus/homology_menu.shtml
Rat Genome Database	http://www.rgd.mcw.edu/VCMAP/mapview.shtml

Quantitative traits analysis

The Jackson Laboratory, Churchill Laboratory	http://www.jax.org/staff/churchill/labsite
R/qtl	http://www.biostat.jhsph.edu/~kbroman/qtl
Web/QTL	http://www.genenetwork.org/home.html
The Jackson Laboratory PGA	http://pga.jax.org/resources/index.html
The Complex Trait Consortium	http://www.complextrait.org

* Members are: BayGenomics (USA), Centre for Modelling Human Disease (Toronto, Canada), Embryonic Stem Cell Database (University of Manitoba, Canada), Exchangeable Gene Trap Clones (Kumamoto University, Japan), German Gene Trap Consortium (Germany), Sanger Institute Gene Trap Resource (Cambridge, UK), Soriano Lab Gene Trap Database (Fred Hutchinson Cancer Research Center, Seattle, USA), TIGEM-IRBM Gene Trap (Naples, Italy). ‡ Database of Cre-expressing strains. GNF, Genomics Institute of the Novartis Research Foundation; NCBI, National Center for Biotechnology information; UCSC, University of California at Santa Cruz.

All these resources have converged to meet the current challenge of complex trait analysis. The hundreds of genetically and phenotypically diverse inbred strains are the starting point from which to explore causal genotype–phenotype relationships and identify QTLs for any measurable trait. Inbred strains carrying spontaneous[5] or engineered[6] mutations allow us to detect QTLs that modify specific disease states. For example, the *Mom1* (modifier of *Min*) gene significantly influences the multiple intestinal neoplasia phenotype that arises in *Min*$^{+/-}$ mice due to defects in the *Apc* (adenomatosis polyposis coli) gene[7]. Derivatives of the inbred strains, such as congenic and recombinant inbred strains, provide resources to identify QTLs, confirm QTL effects and hone in on the causative underlying gene, which can be confirmed with knockout mice[8]. Last, whole-genome sequencing and haplotype structure have led directly to new statistical and bioinformatic tools for QTL studies, and searchable databases are required in all phases of QTL analysis. So, 100 years of resource generation and preservation serves us well today.

Mouse resources

Inbred strains, spontaneous mutations and beyond. Inbred mouse strains are derived from a single parental mating pair with subsequent repeated brother–sister intercrosses and no breeding from non-sibling mice. After 20 generations of inbreeding, the mice are genetically identical and homozygous at all loci. There are hundreds of inbred strains representing a wealth of genetic and phenotypic diversity, with much of the latter still to be examined. New strategies to systematically collect phenotypic data on the common inbred strains from a range of disciplines will tap into this diversity[9–12]. Such a catalogue of phenotypic information allows investigators to choose strains (a) as progenitors in crosses to identify QTLs; (b) as starting points for further model development; (c) for physiological testing and/or drug development; and (d) as sensitized strains for mutation screens.

Following the early emphasis on cancer and transplantation, spontaneous mutations in inbred strains (coisogenic strains[13]) provided researchers entry into normal biological pathways and the pathogenesis of human disease. Heritability testing and linkage mapping established unequivocally the inherited basis of various abnormal phenotypes, and detailed phenotypic analyses revealed physiological mechanisms. For example, long before the obese (*ob*) and diabetes (*db*) gene mutations were identified[14–16], physiological studies revealed that the *ob* gene product suppressed hunger, and required the *db* gene product to do so. Phenotypic analysis of the erythroid ankyrin (ANK1)-deficient normoblastosis (*nb*) anaemic mouse showed age-related ataxia due to Purkinje cell degeneration[17]. This lead to the realization that *Ank1* expression occurred in the brain, particularly the cerebellar Purkinje cells, as well as erythroid tissues, and explained why haemolytic anaemia in humans was sometimes accompanied by neurological defects[18,19].

Mouse models are also invaluable in hypothesis testing. It was predicted that a 440 kDa isoform of another member of the ankyrin gene family, *Ank2*, was specifically targeted to premyelinated axons, and subsequently disappeared as myelination proceeded. In myelin-deficient *shiverer* mice, levels of 440-ANK2 were increased, supporting the hypothesis[20]. Even in the post-genomic era, mouse models are crucial to gene discovery. Phenotyping and genetic mapping of the spontaneous anaemia mutation *nm1054* (REF. 21), for example, led to the unexpected discovery of a role for the *Steap3* (six-transmembrane epithelial antigen of the prostate 3) gene product in iron uptake in red cells[22].

The analysis of spontaneous mutations is therefore a time-honoured method of gene discovery and elucidation of gene function. However, the spontaneous mutation rate is relatively low. Mutation frequency can be accelerated using chemical mutagens, primarily *N*-ethyl-*N*-nitrosourea (ENU)[23–29]. Mouse mutagenesis can provide novel mutants in new phenotypic areas, hypomorphic alleles at known loci, and even a series of mutant alleles within the same locus[30]. In response to the initiatives of the US National Institutes of Health (NIH) to further the use of the mouse in biomedical research (see Trans-NIH Mouse Initiatives), several large-scale mutagenesis programmes, each screening for specific phenotypes (TABLE 1), are now operating.

Derivatives of inbred strains. Specialized strains (FIG. 1) shuffle the genome and increase the diversity that is available to researchers. Recombinant inbred strains are developed by crossing two different parental strains and then intercrossing their offspring for 20 generations[31,32] (FIG. 1a). Each recombinant inbred strain has a unique, fixed combination of the original parental genomes. An example is the reciprocal AXB, BXA strains, which are derived from the A/J and C57BL/6 (B6) inbred strains. Recombinant inbred strains were used in gene mapping before the development of short sequence length polymorphic markers[33] (SSLP markers), and are now important in complex trait analyses. However, as the number of strains in these classical recombinant inbred strain sets is relatively small (for example, the original BXD recombinant inbred set consists of 26 strains), genetic and phenotypic diversity, and therefore QTL detection and resolution, are limited. Recently, expanded recombinant inbred panels[34,35] and advanced intercross lines (AIL), in which generations beyond F_2 are generated to increase recombination frequency[36] (FIG. 1b), have increased the power of QTL detection (reviewed in REF. 37). A variation of the recombinant inbred strain concept is the Collaborative Cross, proposed by the Complex Trait Consortium[8], which can generate novel and diverse allelic combinations (described below). Another resource that is used for QTL mapping is heterogeneous stock mice, which originate from an eight-way cross, but are kept heterozygous by random mating[38].

In congenic strains — first introduced by George Snell in his Nobel-Prize-winning research on histocompatibility loci[3] — a single locus (and surrounding DNA) from one strain is moved to another strain by repeated backcrossing and intercrossing (FIG. 1c). The amount of the donor genome in the new strain depends on the

Coisogenic strain
Strains of mice that differ at a single locus. When a spontaneous mutation occurs in an inbred strain, the mutant and the non-mutant mice are coisogenic.

Purkinje cell
Large neurons with highly branched dendritic trees; Purkinje cells provide the only neural output from the cerebellum.

Myelination
The process of the formation of the myelin sheaths of axons.

Short sequence length polymorphic markers
Tandem repeats, usually of 2–3 bp; the number of repeats varies (is polymorphic) depending on the strain.

Table 1 | **Mutagenesis centres**

Programme	Phenotypes*	Web URL
Mouse Heart, Lung, Blood and Sleep Disorders, The Jackson Laboratory	Hematopoiesis, glucose, plasma lipids, sleep, lung, hypertension, obesity, kidney	http://pga.jax.org/index.html
Neuroscience Mutagenesis Facility, The Jackson Laboratory	Gait, vision, seizure threshold, hearing, epilepsy	http://nmf.jax.org/index.html
Reproductive Genomics, The Jackson Laboratory	Sperm count and motility, oocyte abundance, development	http://reprogenomics.jax.org/index.html
The Neurogenomics Project at Northwestern University	Learning and memory, vision, circadian rhythm, psychostimulant response	http://genome.northwestern.edu/neuro
The Neuromutagenesis Project, Tennessee Mouse Genome Consortium	Aggression, alcohol and drug abuse, hearing, vision, behaviour, ageing	http://www.tnmouse.org/neuromutagenesis/index.html
Oak Ridge National Laboratory's Mouse Genetics and Genomics Program	Urine, blood, behaviour, embryonic lethal	http://bio.lsd.ornl.gov/mgd
Mouse Mutagenesis Center for Developmental Defects at Baylor College of Medicine	Development, haematopoiesis, bone, urogenital	http://www.mouse-genome.bcm.tmc.edu/ENU/MutagenesisProj.asp
The Harwell ENU-Mutagenesis Programme	Neurological	http://www.mgu.har.mrc.ac.uk/facilities/mutagenesis/mutabase
Toronto Center for Phenogenomics	Cardiovascular, renal, skeletal, metabolic	http://cmhd.mshri.on.ca/
McLaughlin Research Institute Mutagenesis Project	Prion disorders, Alzheimer disease	http://www.montana.edu/wwwmri/enump.html
GSF, Germany	Allergy, immunology, dysmorphology	http://www.gsf.de/ieg/groups/genome/enu/mutants.html

*Not intended to be comprehensive; see web sites for complete details.

number of backcross generations[39]. Congenic strains are particularly useful, as they control for phenotypic differences due to strain background. Polymorphic DNA markers allow accelerated creation of congenic strains by marker-assisted breeding ('speed' congenics)[40].

Variations on the congenic-strain concept are chromosome substitution strains (CSS) and genome tagged mice (GTM) (FIG. 1d,e). In GTM, large overlapping segments of each chromosome are transferred to a different genetic background and bred to homozygosity[41]. The 65 B6.DBA/2 congenic strains contain overlapping portions of each DBA/2 autosome, with mapped boundaries, on an otherwise B6 background[41,42]. A similar set of B6.CAST/Ei GTM are also available[42] (see the UCLA GTM Resource for information on both sets). GTM allow one to confirm the presence of a QTL on a given segment, or even to subdivide a QTL interval. Similarly, CSS are available in which a complete single chromosome from a donor strain is introgressed onto another strain. The first CSS set consisted of A/J strain chromosomes on a B6 strain background[43]. A second set with PWD/Ph strain chromosomes on the B6 background was recently constructed. Both sets are available through The Jackson Laboratory. Another CSS set is being constructed with the DBA/2 strain as the background and the 'extremely big' DU6i strain as the donor[44]. These strains allow for rapid mapping of phenotypic traits to a single chromosome, or to subsets of chromosomes[43,45].

Designer mutants. Transgenic mice[46] and homologous recombination in embryonic stem (ES) cells[47,48] provided the tools to manipulate mammalian genes *in vivo*. Unlike classical phenotype-driven strategies (for example, positional cloning of a spontaneous mutation), manipulation of the mouse genome requires previous knowledge of gene structure, which is now provided by the complete mouse sequence. Temporal and tissue-specific manipulation have added to the enormous contributions of genetically engineered mice to our understanding of mammalian gene function (see REF. 6 for the history and current state of transgenic technologies, and REF. 49 for an excellent practical review on conditional targeting strategies).

There are now several large-scale knockout projects, including the publicly funded gene-trap efforts that are being coordinated by the International Gene Trap Consortium (BOX 1), and the private efforts of Lexicon Genetics and Deltagen. However, it is estimated that only ~10% of the known mouse genes have been knocked out[50], and most of these are not freely available to researchers. For this reason, the NIH initiated the Knockout Mouse Project (KOMP)[50] to generate a public repository of null ES cell lines representing every gene in the mouse genome. International partners in this effort are the North American Conditional Mouse Mutagenesis Project (NorCOMM) and the European Conditional Mouse Mutagenesis Program (EUCOMM), both of which focus on generating conditional alleles. Furthermore, NIH has contracted with Lexicon and Deltagen to make 251 of their previously generated knockout mouse lines freely available (see Deltagen and Lexicon Knockout Mice and Phenotypic Data), and have initiated a repatriation project to place existing knockout mice in public repositories to facilitate their distribution. Searchable databases to gain information, and to obtain mouse knockout and transgenic lines and vectors for generating lines, are listed in BOX 1.

Figure 1 | **Derivatives of inbred strains. a** | Recombinant inbred strains are developed by crossing two different inbred parental strains to produce F_1 offspring (obligate heterozygotes at all loci). From there, a series of brother–sister matings are established, and their offspring are repeatedly intercrossed for at least 20 generations. This produces fully inbred strains, each of which is homozygous at all loci for a unique combination of the original parental genomes. **b** | In advanced intercross lines (AIL), the goal is to increase recombination frequency, so matings between siblings and cousins are avoided. By providing large numbers of animals that carry many additional genetic breakpoints, AILs are particularly useful in narrowing QTL confidence intervals. **c** | Congenic strains are produced with the goal to transfer a single locus such as a mutant gene from one genetic background to another. In this example, a chromosome 6 (Chr 6) locus is illustrated. The mouse carrying the locus to be transferred is mated, or 'outcrossed,' to the strain of choice to produce obligate heterozygotes. The heterozygotes are then intercrossed, and the process of outcrossing and intercrossing, with selection for the locus of interest at all outcross generations, is repeated. In practice, a single locus alone can never be transferred to a second strain; some flanking as well as non-flanking DNA is carried with it. The amount can be minimized by increasing the number of outcross generations[19]. **d** | In chromosome substitution strains (CSS, formerly called consomic strains), one chromosome in its entirety is transferred from one strain background to another. **e** | Genome tagged mice (GTM) are similar in concept to a congenic strain, but the idea is to not only transfer a single locus to another genetic background, but to transfer large, overlapping regions of each chromosome from one strain to another, and to build up a collection of such strains that covers the whole genome.

The mouse genome and phenome

Genome. Positional cloning became much easier with the publication of the mouse genome sequence[2], as annotated sequences within crucial chromosomal intervals could now be accessed. Tedious physical mapping with yeast and BACs was no longer required, nor the labour-intensive search for polymorphic markers. Moreover, enormous sequence variation was shown to exist among the inbred strains that were sequenced (A/J, DBA/2J, 129 and B6), which was predicted to make QTL gene identification easier[51]. Identifying the gene underlying a QTL remains a challenging task, but combined use of mouse resources and newly emerging bioinformatics and statistical tools is making significant inroads[52].

Building a catalogue of genome annotations is just the beginning for biology in the post-genomic era[53,54]. To derive new insights into fundamental biological processes will require understanding how genome features interact in pathways and networks in the cell, and how perturbations contribute to disease[53,55]. Powerful

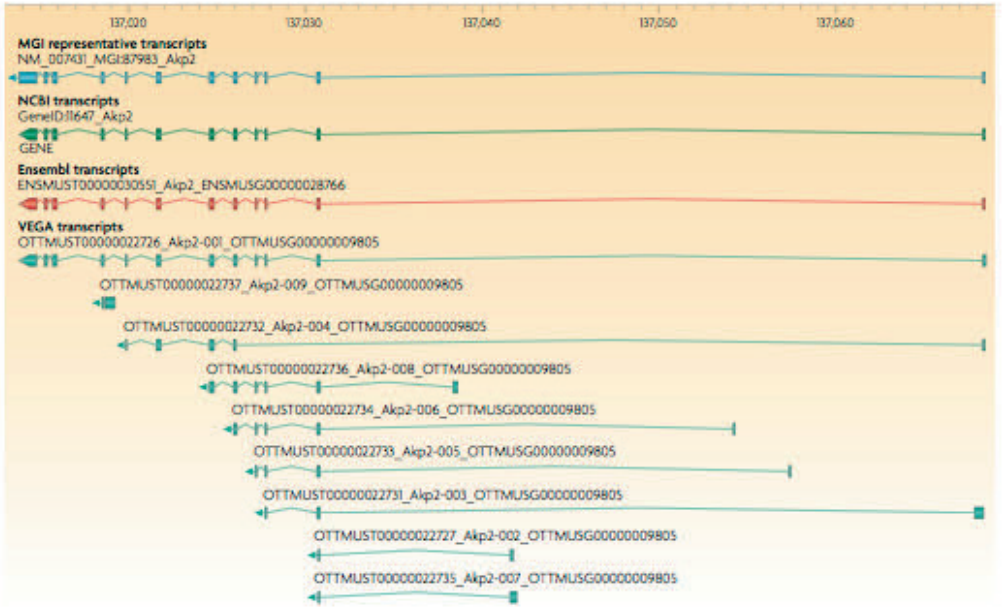

Figure 2 | **Search results for *Akp2* in the Mouse Genome Informatics (MGI) Mouse GBrowse.** The annotations all relate to the same gene — *Akp2* (chr4:137013809..137068460; Build 36) — but there are significant differences in the structural details between VEGA (Vertebrate Genome Annotation Group) manual annotation, showing multiple transcripts, and the computational gene predictions of NCBI (National Center for Biotechnology Information) and Ensembl. The scale shows the base position (kb) along the chromosome.

web-based programs are now available for pathway analysis [BOX 1].

Access to the annotated mouse genome sequence and the integration of genome annotations with biological knowledge is supported by several informatics resources, such as the Ensembl group at the Sanger Institute[56], the Map Viewer at the US National Center for Biotechnology Information (NCBI), and the Genome Browser at the University of California at Santa Cruz[57,58]. The Mouse Genome Informatics (MGI) database, the community model organism database for the laboratory mouse[59,60], integrates genome annotations with biological knowledge about genes and gene products, including the associations between genes and phenotypes, SNPs, mammalian homology, gene expression and functional (that is, gene ontology) classifications[61] [BOX 1].

Researchers who use online mouse genome resources must keep in mind that mouse genome-sequence reference assembly, computational gene predictions and biological knowledge about mouse genes and gene products change over time. Each new version of the genome assembly results in new gene predictions that are generated by computational genome annotation pipelines at NCBI and Ensembl. The Vertebrate Genome Annotation Group (VEGA) at the Sanger Institute has undertaken the Herculean task of manually evaluating all annotations[62]. Importantly, gene predictions, base positions and transcript information will differ between databases, and one should not rely on any single resource. The MGI database group evaluates annotations to determine equivalency or novelty [FIG. 2] and relies on expert curators to associate these genes with biological knowledge.

Within a database, base positions change from one assembly to the next, SNP reference numbers change and information about genes evolves. Attention to these details is required when studies overlap the emergence of a new assembly. Genome browsers back-reference previous assemblies (although not all of them), allowing one to update information. For example, the peak LOD score position and the 95% confidence interval for each QTL that is identified will invariably change, although usually only slightly.

Phenome. In 1999, the Strain Characterization Project, now called the Mouse Phenome Project, was conceived at a meeting of physiologists, behaviourists, geneticists and bioinformaticians at The Jackson Laboratory[11]. The group realized that a publicly available, systematic survey of phenotypic and genotypic information on a standard set of inbred mouse strains would be invaluable. The

LOD score
The 'Logarithm of odds' score; the base-10 logarithm of a likelihood ratio (the odds), which is often used in the context of genetic mapping to indicate significance thresholds.

resultant Mouse Phenome Database (MPD) currently lists 65 different projects, encompassing a broad range of biological disciplines, for up to 48 inbred strains[8,63,64]. It also supports basic statistical analyses and comparisons of data across or within strains and studies. This resource is indispensable for association-mapping experiments, and for identifying mouse strains that display desired phenotypic characteristics, such as the most appropriate models of human disease[9,63,66]. A recently added tool — 'Find mouse models' — allows investigators to search for strains that display combinations of phenotypes. Phenotyping efforts by the German Mouse Clinic[67,68] and Eumorphia EMPReSS[12] (BOX 1) have also been instituted. These differ from the MPD in that they emphasize standardized phenotyping protocols. Detailed procedures for assaying a range of basic and sophisticated phenotypes are available.

Genetic and phenotypic diversity among inbred strains. With the recognition that the vast majority of human disease results from deleterious combinations of normal polymorphic alleles, complex trait analysis in animal models becomes paramount.

The first extensive sets of polymorphic markers were SSLP[33], and other anonymous DNA markers such as expressed sequence tags (ESTs). These were useful in gene mapping, and led to considerable saturation of linkage maps[69]. Furthermore, dense SSLP maps made positional cloning and candidate-gene analysis feasible. Backcrosses or intercrosses between extensively typed inbred strains carrying single-gene mutations and a second, highly polymorphic strain allowed one to track the segregation of SSLP markers with a mutant phenotype. Today, SNPs, which represent natural sequence variation, are the markers of choice in genetic mapping[70–72], and SNPs that vary between more than 100 inbred strains have been identified. The US National Institute of Environmental Health Sciences (NIEHS)-sponsored resequencing of 15 mouse strains has identified millions of SNPs (see the Perlegen Sciences web site), and many other efforts are also under way (BOX 1). NCBI's dbSNP is the primary public repository for mouse SNP data and is updated regularly. These data are downloaded regularly to databases such as MGI and MPD, where they are integrated with other biological data and can be accessed using query and visualization tools. Many of the main providers of SNP data (for example, Roche and the Genomics Institute of the Novartis Research Foundation (GNF)) also provide access to these data from their own web sites.

Integrating resources in QTL analysis

QTL analysis in mice and human complex disease. The fact that human and mouse genes are arranged syntenically[73] makes it possible to cross-identify genes of interest. QTL studies for a given trait in mice and humans have found that the corresponding QTLs are located in homologous regions. This phenomenon, termed concordance, was first noted for hypertension[74], and subsequently for atherosclerosis[75], bone density[76], kidney disease[77], plasma lipids and asthma[78,79]. Of human plasma-lipid QTLs, 93% of those for HDL cholesterol, 100% for low density lipoprotein (LDL) cholesterol and 80% for triglycerides have corresponding QTLs in the mouse genome[79]. The predictive value of QTL identification in humans by their identification in mice is therefore readily apparent. Moreover, in cases in which the causal gene that underlies a QTL for a disease phenotype in mice has been identified, the same gene can be seen to underlie the corresponding QTL in humans[80,81]. Recent examples include the *Hc* gene (complement factor 5) for liver fibrosis[81], *Tnfsf4* (Ox40 ligand) for atherosclerosis[80], and *Ctla4* for type I diabetes[82].

According to the MGI database (BOX 1) there are more than 2,900 published QTL studies in the mouse[37], but few (~20) genes underlying QTLs have been identified[37]. Some are more optimistic[83] than others[37] on the prospects of improving this, but recent advances are expediting the process. These include: analysis of conserved chromosomes across species; the use of databases (sequence, expression and phenome databases); CSS, GTM and AIL; statistical methods; and SNP analysis (haplotyping). These methods, especially when used together, can narrow QTL intervals to give a manageable number of candidate genes[52].

The most dramatic advances in QTL analysis derive from resequencing efforts. The haplotype structure of the laboratory mouse genome, first described by Wade *et al.*[84], is a mosaic of segments that are derived primarily from two ancestral sources, *Mus musculus musculus* and *Mus musculus domesticus* (with a minor contribution from *Mus musculus castaneus*)[71,85]. Approximately 97% of the variation between inbred strains is ancestral, reflecting the fact that they shared a common ancestor only ~100 years ago. Regions that are identical between two strains that are used to detect a QTL — that is, regions that are 'identical by descent' — are highly unlikely to contain causal genetic polymorphisms underlying the QTL[86]. A caveat is that if the causal mutation occurred after the strains were separated, the underlying gene could indeed reside in a region of identity, but this is rare. These observations have practical implications in QTL analysis.

Identifying QTLs: practical considerations. A QTL study initially involves selection of appropriate parental inbred strains, establishment of a backcross or F_2 intercross, phenotyping and genotyping of progeny for polymorphic markers that are evenly spaced throughout the genome, and regression analysis[87]. The goal is to identify genomic regions in which allelic variation is associated with phenotypic variation. For detecting modifiers of a specific disease phenotype, one of the parental strains should carry a disease-causing allele[7,88,89] (see REF. 87 for a discussion of the statistical paradigm for QTL analysis, including genome scans to detect single (main-effect) QTLs, permutation testing for significance thresholds[90], calculation of confidence intervals, pairwise comparisons to detect QTL–QTL interactions and epistasis, generation of multi-QTL models, and determination of allele effects).

Synteny
Genes that occur in the same order in different species; the chromosomes of the species are then said to be syntenic in those regions.

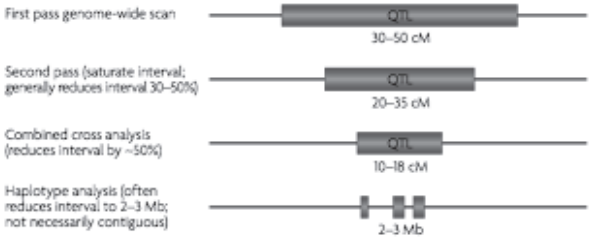

Figure 3 | **Identification of QTL and narrowing QTL intervals.** Following an initial 'first pass' genome-wide scan of 200–300 F_2 progeny (a typical starting point), QTL confidence intervals are usually large, spanning as much as 30–50 cM. One strategy to reduce the size of the interval is to saturate the QTL interval with additional polymorphic markers and add more progeny to the analysis. This strategy often decreases the interval by ~30–50%. Clearly, however, such an interval is still likely to contain too many genes and transcripts for meaningful candidate-gene analysis. A highly effective strategy is to establish more than one cross, allowing the application of powerful new statistical (combined cross analysis) and bioinformatics (haplotype analysis) tools to markedly reduce QTL intervals in silico, that is, without further 'wet lab' work.

The MPD provides an invaluable starting point, allowing one to choose parental strains for QTL crosses. The greater the statistical difference for the quantitative trait is between strains, the greater the likelihood of detecting significant QTLs[91]. Clearly, the more phenotyping data that is deposited into the MPD, the greater its utility. For example, we sought to identify QTLs that influence steady-state peripheral blood cell traits[92,93]. Baseline peripheral blood counts are significant risk factors for cardiovascular disease and stroke in the general population, and for disease severity in sickle cell anaemia[94–102]. We chose parental strains NZW/LacJ (NZ), C57BLKS/J (KS) and SM/J (SM) because of their marked differences in these parameters, particularly the white blood cell (WBC) count. In two F_2 intercrosses, NZ × SM and KS × SM, we detected multiple significant QTLs for several different blood traits using just 186 progeny[92,93]. Following this initial genome-wide scan, QTL intervals typically span large chromosomal segments. For example, of the six QTLs identified for WBC count, the smallest interval was 14 cM, or ~28 Mb[93] (often 95% confidence intervals span 30 cM or more). Candidate-gene analysis in an interval of this size is usually impractical. Increasing the number of F_2 animals that are analysed, even by several hundred, will not decrease the interval to a manageable range[103], nor does saturating the QTL interval with additional markers, in our experience[93] (FIG. 3).

Narrowing QTL intervals. GTM and advanced intercross lines can be used to significantly narrow QTL intervals and resolve linked QTLs (FIG. 4a,b). Homology data can also be an effective tool[104–106]. If a human and a rodent QTL for the same phenotype have been found in a conserved (syntenic) location, one can use homology relationships to define the overlapping genomic segments, and thereby narrow the interval. However, for many phenotypes, QTLs have not been identified in multiple species.

The use of mouse-strain resources to narrow QTL intervals requires a significant investment of time, animal space and money. So, statistical and bioinformatic resources are essential alternatives. Combining data from multiple QTL crosses, in which overlapping QTLs have been identified, increases statistical power[107]. The more crosses that are combined, the more power is achieved. The phenotypic and genotypic information from multiple crosses is combined into a single QTL regression analysis. As most genetic variation between mouse strains is derived primarily from *M. m. musculus* and *M. m. domesticus*[66,70–72,85], combined cross analysis assumes a common biallelic mode of inheritance for each locus. All strains carrying an allele that results in a high value for a given trait are assumed to share the same ancestral allele for the underlying causal gene, whereas strains carrying an allele that produces low values for the trait are assumed to also share the same ancestral allele, which is necessarily different from the high-scoring allele.

The main assumption that underlies this combined cross analysis is that the same causal gene underlies overlapping QTLs that are identified in different crosses. If this assumption is correct, the increased numbers of progeny and markers add statistical power to the analysis by providing more recombination, allowing one to detect QTLs with smaller effects and narrow confidence intervals. Even if the assumption is incorrect, the analysis is still powerful, as it often resolves closely linked QTL into distinct peaks. In analysing peripheral blood traits, we identified a common QTL on chromosome 7 for the red cell mean corpuscular volume (MCV), *Mcvq1* (REF. 93). In both crosses, the 95% confidence interval was 18 cM, but combining them reduced it to 12 cM (by 33%), despite the fact that the analysis involved just two crosses and three strains (FIG. 4c). This decreased the number of genes and transcripts within the interval from ~300 to 110 (using Ensembl, Build 35). In other studies, combined cross analysis decreased a QTL for blood pressure on chromosome 1 from 42 cM to 18 cM[108], a chromosome 4 HDL QTL was decreased from 30 cM to 10 cM[107], and a bone density QTL on chromosome 7 was decreased from 34 cM to 22 cM[109]. In all cases, without carrying out any further wet-lab studies, the confidence interval was significantly reduced.

Haplotype analysis can decrease confidence intervals even more dramatically, by quickly identifying high-priority regions within a QTL interval that are likely to contain the causal polymorphism; that is, regions that are not identical by descent[108]. To be most effective, a dense SNP map is required — another strong consideration in the choice of parental strains. For the 15 strains that are being resequenced by NIEHS, and for the reference strain, B6, SNP density is extremely high. As with combined cross analysis, multiple crosses between different strains greatly enhance power[110]. Haplotype analysis of a mean platelet volume (MPV) QTL, *Mpvq1* (REF. 93), showed that high-MPV strains KS and NZ shared a haplotype block that consists of seven contiguous SNPs within the *Mpvq1* 95% confidence interval on chromosome 15, which differed from that of the low MPV strain, SM. This narrowed the *Mpvq1* interval from 2 cM (~4 Mb) to just

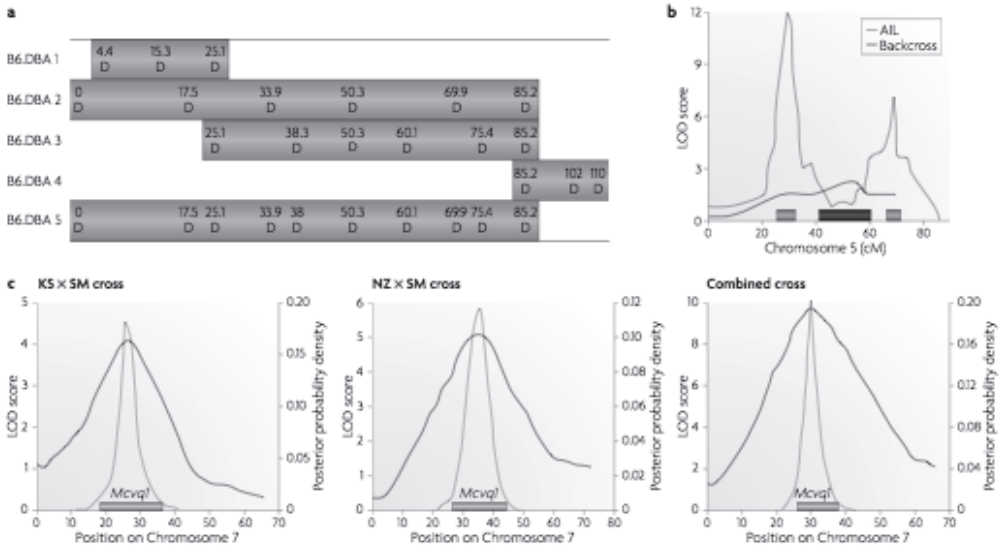

Figure 4 | **Strategies for narrowing QTL confidence intervals. a** | Genome tagged mice (GTM). The five chromosome 1 B6.DBA GTM strains are shown. Each row represents a strain, and each strain contains a unique segment of the DBA chromosome 1 (shown in green) on the B6 background. The numbers above indicate the chromosomal regions, expressed in cM. These strains can be phenotyped for traits in which QTLs on chromosome 1 are detected; the boundaries between the DBA and B6 chromosomal segments, together with phenotypic information, can narrow the QTL interval. For example, a QTL that spans the region from 75 to 100 cM can potentially be subdivided using strains 3 and 4. **b** | Advanced intercross lines (AIL). AILs can both significantly narrow QTLs and resolve closely linked QTLs. Mice are generated for multiple generations beyond F_2, taking care to avoid mating siblings and cousins. Recombinations accumulate with each successive generation. Closely spaced markers within a previously identified QTL interval are then typed. Wang *et al.*[128] resolved a broad QTL for high density lipoprotein (HDL) on chromosome 5 into two distinct QTLs by analysing AIL progeny. Note that the 95% confidence interval that was calculated for the backcross (shown as a black rectangle) is ~20 cM, and does not correspond to the confidence interval that was resolved by AIL analysis (shown as red rectangles). **c** | Combined cross analysis. An example of combined cross analysis for the Mcvq1 locus is shown. Combining data from two crosses (K5 × SM and NZ × SM) decreases the Mcvq1 confidence interval (shown as a blue box)[92]. The left scale and black lines indicate the logarithm of odds (LOD) score; the right scale and blue lines indicate the posterior probability density[92]. The posterior probability density defines the 95% confidence interval. It gives results that are similar to the more familiar 1.5 'LOD support interval,' but is better justified on theoretical grounds. Part **b** is modified with permission from REF. 128 © (2003) Cold Spring Harbor Laboratory Press. Part **c** is modified with permission from REF. 92 © (2006) Springer Verlag.

0.35 Mb. Moreover, the analysis revealed a strong candidate gene, *Nfe2*, which encodes a transcription factor that is required for pro-platelet formation[111–116]; the rate of platelet production influences platelet size[117].

Expression databases are invaluable in QTL candidate-gene analysis. Microarray databases such as SymAtlas can confirm that a candidate gene is appropriately expressed. The Gene Expression Omnibus (GEO) and The Institute for Genome Research (TIGR) databases (BOX 1) allow identification of genes that are differentially expressed among inbred strains. Two specialty expression databases are GenSat, which provides gene-expression information for the mouse central nervous system obtained from BAC transgenics, and the Allen Institute Brain Atlas (BOX1).

eQTLs. Gene expression levels can be considered a quantitative trait[118]. Using microarray data from tissues of F_2 animals, one can obtain expression QTLs (eQTLs) that regulate transcript levels[119,120]. eQTLs that co-localize with the gene in which transcription is being regulated are referred to as *cis* eQTLs. When *cis* eQTLs co-localize with QTLs for a clinical trait, they can greatly facilitate the search for causal genes[121,122]. However, microarrays and eQTLs have an even greater utility in identifying pathways, because one can interrogate many transcripts simultaneously. Integrating genetic and gene-expression data on a global basis might allow the identification of key hub regulatory genes that have a strong influence on a clinical trait[119,121].

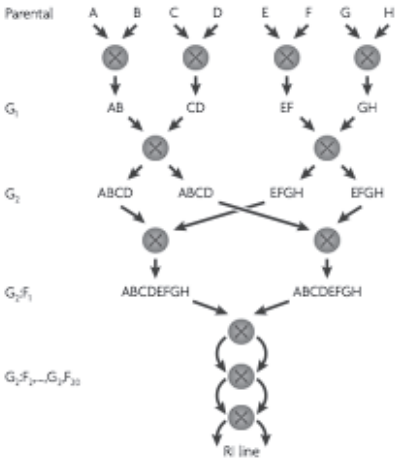

Figure 5 | **The Collaborative Cross.** The eight parental strains (shown as strains A–H) are C57BL/6J, A/J, CAST/Ei, NOD/LtJ, NZO, WSB/Ei, PWK/Ph and 129S1/SvImJ. A total of 56 (28 reciprocal) G_1s will be produced (diallele cross). These will be intercrossed to produce 1680 pairs that do not share parental strains. Two further generations of crossing before repeated inbreeding ensure that each recombinant strain is descended from all eight parental strains. The Collaborative Cross is a community effort involving several groups, including The Jackson Laboratory, Oak Ridge National Laboratory and the Wellcome Trust. Eight-way crosses for more than 500 lines are in progress at these facilities.

Complex trait analysis: a community effort. Clearly, using data from multiple crosses allows for the statistical analysis of combined cross data and enhances haplotype analysis. However, the step of combining crosses statistically requires access to the raw data; published information is not sufficient. The QTL Archive at The Jackson Laboratory is a growing resource that provides access to raw data from QTL studies using rodent crosses. A community data submittal link is included and submissions by the research community are encouraged. Sharing raw QTL cross data not only protects that data from loss, but allows new advances in QTL analysis to be applied to previously generated crosses.

The Collaborative Cross; a vision for the future. The Complex Trait Consortium has initiated a community effort to revolutionize the study of complex diseases. The Collaborative Cross will provide a common, genetically defined and reproducible panel of strains that are available at a reasonable cost[123,124]. It will start with eight genetically diverse inbred strains (including wild-type-derived strains) and end with 1,000 recombinant inbred strains (FIG. 5). All recombination events will be mapped in each strain and extensive genotyping will be carried out. Because all the strains are fully reproducible, genotyping need only be done once. Data on these strains can be integrated across laboratories that carry out diverse and specialized phenotyping procedures.

Many recombinant inbred strain sets exist already, so what will the Collaborative Cross add? It is estimated that each strain will possess 135 unique recombination events, and that the Collaborative Cross will capture sequence polymorphisms at 100–200 bp intervals[124]. Therefore, the genetic diversity among 1,000 strains will be monumental. QTL intervals of ~0.1 cM (~200,000 bp) are possible. By intercrossing recombinant inbred strains to produce F_1 progeny, even more genetic and phenotypic diversity is produced. Notably, recombinant inbred F_1 progeny will have reproducible genotypes and low inbreeding coefficients, making their levels of heterozygosity more representative of human populations[124].

Many of the considerations that went into the final design of the Collaborative Cross are purely practical, designed to have the finished lines available within about 5 years at a reasonable cost. For example, intercrossing F_1 progeny for two or more generations would increase the number of recombinations, but simulations indicate that the gain would be minimal. Likewise, although a recent simulation study indicates other systems might achieve similar power and resolution[123], their construction is also more costly.

Finally, can the average laboratory phenotype 1,000 recombinant inbred strains? With relatively simple high-throughput assays, phenotyping might not be an obstacle, although it could be for more sophisticated testing. Fortunately, one can use a subset (~100–200) of the collaborative recombinant inbred strains to identify the main QTLs, and then carry out further phenotyping using 100–200 other strains that have recombinations within the QTL interval. Extensive information on this strategy and all other aspects of the Collaborative Cross are available at the Complex Trait Consortium web site (BOX 1).

Future directions

New bioinformatic and statistical approaches to the analysis of complex traits continue to emerge, all of which depend on highly dense genetic maps and extensive phenotyping data. An example is *in silico* QTL analysis, now known as haplotype association mapping (HAM)[65]. Grupe *et al.*[125] first developed a computer algorithm to predict chromosomal regions that regulate phenotypic traits among inbred strains. *In silico* mapping is most robust when large numbers of markers and phenotyped strains are available, although the exact numbers that are required and how to best analyse the data statistically remain matters of discussion. In our experience, *in silico* analyses detect large numbers of potential loci, and it is problematic to distinguish the genuine ones. However, when overlap occurs with the loci that are found in experimental studies[66,125,126], the *in-silico*-detected QTLs substantially reduce the confidence interval[52].

A looming challenge for complex trait analysis is to avoid simply cataloguing the genes and interactions that produce a phenotype, but actually elucidate the causal relationships among them. Structural equation modelling is a statistical method that can infer the 'architecture' of such networks[127]; that is, the magnitude and direction

of causal relationships among interrelated phenotypes. New methods such as HAM and structural modelling must be accessible to the average member of the scientific community. This will require web-based interfaces to facilitate the use of sophisticated statistical programs. One example is a new software package for QTL analysis, J/qtl, a JAVA-based package that provides an interface to the powerful R/qtl package, and simplifies the analysis of QTL data for the average laboratory. Likewise, the Genome Variation Server allows easy access to the human genotype data that is found in dbSNP and at the International HapMap Project web site, and provides tools for analysis of genotype data. Last, data and resources in rat genetics are also freely available from PhysGen. Notably, all of these examples are part of the National Heart, Lung and Blood Institute's 'Programs for Genomic Applications' (PGA), where the goal is to place genomic resources and tools into the hands of investigators everywhere. Similar efforts by all, including individual investigators, to make resources and data accessible will be of crucial importance in advancing complex trait analysis and the dissection of human disease.

1. Little, C. C. US science wars against an unknown enemy: cancer. *Life* **2**, 11–17 (1957).
2. Waterston, R. H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
A description and analysis of the initial mouse genome sequence.
3. Paigen, K. One hundred years of mouse genetics: an intellectual history. I. The classical period (1902–1980). *Genetics* **163**, 1–7 (2003).
4. Paigen, K. One hundred years of mouse genetics: an intellectual history. II. The molecular revolution (1981–2002). *Genetics* **163**, 1227–1235 (2003).
References 3 and 4 give an insightful history of the past 100 years of mouse genetics.
5. Doolittle, D. P., Davisson, M. T., Guidi, J. N. & Green, M. C. in *Genetic Variants and Strains of the Laboratory Mouse* Vol. 1 (eds Lyon, M. F., Rastan, S. & Brown, S. D. M.) 817–854 (Oxford Univ. Press, New York, 1996).
6. Lee, D. & Threadgill, D. W. Investigating gene function using mouse models. *Curr. Opin. Genet. Dev.* **14**, 246–252 (2004).
7. Dietrich, W. F. *et al.* Genetic identification of *Mom-1*, a major modifier locus affecting Min-induced intestinal neoplasia in the mouse. *Cell* **75**, 631–639 (1993).
8. Abiola, O. *et al.* The nature and identification of quantitative trait loci: a community's view. *Nature Rev. Genet.* **4**, 911–916 (2003).
9. Bogue, M. A. & Grubb, S. C. The mouse phenome project. *Genetica* **122**, 71–74 (2004).
10. Grubb, S. C., Churchill, G. A. & Bogue, M. A. A collaborative database of inbred mouse strain characteristics. *Bioinformatics* **20**, 2857–2859 (2004).
11. Paigen, K. & Eppig, J. T. A mouse phenome project. *Mamm. Genome* **11**, 715–717 (2000).
12. Brown, S. D., Chambon, P. & de Angelis, M. H. EMPReSS: standardized phenotype screens for functional annotation of the mouse genome. *Nature Genet.* **37**, 1155 (2005).
13. Silver, L. M. *Mouse Genetics* 362 (Oxford Univ. Press, New York, 1995).
The standard text on mouse genetics, introducing the history of the laboratory mouse, breeding strategies and other widely used methods.
Available free online (see Further information).
14. Lee, G. H. *et al.* Abnormal splicing of the leptin receptor in diabetic mice. *Nature* **379**, 632–635 (1996).
15. Chen, H. *et al.* Evidence that the diabetes gene encodes the leptin receptor: identification of a mutation in the leptin receptor gene in *db/db* mice. *Cell* **84**, 491–495 (1996).
16. Zhang, Y. *et al.* Positional cloning of the mouse obese gene and its human homologue. *Nature* **372**, 425–432 (1994).
17. Peters, L. L. *et al.* Purkinje cell degeneration associated with erythroid ankyrin deficiency in *nb/nb* mice. *J. Cell Biol.* **144**, 1233–1241 (1999).
18. Bennett, V. Of mice and men: the mice were right. *J. Clin. Invest.* **95**, 921–922 (1995).
19. McCann, S. R. & Jacob, H. S. Spinal cord disease in hereditary spherocytosis: report of two cases with a hypothesized common mechanism for neurologic and red cell abnormalities. *Blood* **48**, 259–263 (1976).
20. Chan, W., Kordeli, E. & Bennett, V. 440-kD ankyrinB: structure of the major developmentally regulated domain and selective localization in unmyelinated axons. *J. Cell Biol.* **123**, 1463–1473 (1993).
21. Ohgami, R. S. *et al.* nm1054: a spontaneous, recessive, hypochromic, microcytic anemia mutation in the mouse. *Blood* **106**, 3625–3631 (2005).
22. Ohgami, R. S. *et al.* Identification of a ferrireductase required for efficient transferrin-dependent iron uptake in erythroid cells. *Nature Genet.* **37**, 1264–1269 (2005).
23. Beier, D. R. ENU mutagenesis: a work in progress. *Physiol. Genomics* **11**, 111–113 (2002).
24. Brown, S. D. M. Mouse models of genetic disease: new approaches, new paradigms. *J. Inherited Metab. Dis.* **21**, 552–559 (1998).
25. Clark, A. T. *et al.* Implementing large-scale ENU mutagenesis screens in North America. *Genetica* **122**, 51–64 (2004).
26. Cordes, S. P. *N*-ethyl-*N*-nitrosourea mutagenesis: boarding the mouse mutant express. *Microbiol. Mol. Biol. Rev.* **69**, 426–439 (2005).
27. de Angelis, M. H. *et al.* Genome-wide, large-scale production of mutant mice by ENU mutagenesis. *Nature Genet.* **25**, 444–447 (2000).
28. Justice, M. J., Noveroske, J. K., Weber, J. S., Zheng, B. & Bradley, A. Mouse ENU mutagenesis. *Hum. Mol. Genet.* **8**, 1955–1963 (1999).
29. Nadeau, J. H. *et al.* Sequence interpretation. Functional annotation of mouse genome sequences. *Science* **291**, 1251–1255 (2001).
30. O'Brien, T. P. in *Systematic Evaluation of the Mouse Eye* (ed. Smith, R. S.) 93–108 (CRC, New York, 2002).
31. Bailey, D. W. Recombinant-inbred strains. An aid to finding identity, linkage, and function of histocompatibility and other genes. *Transplantation* **11**, 325–327 (1971).
32. Swank, R. T. & Bailey, D. W. Recombinant inbred lines: value in the genetic analysis of biochemical variants. *Science* **181**, 1249–1252 (1973).
33. Dietrich, W. *et al.* in *Genomic Maps, Nonhuman Vertebrates* Book 4 (ed. O'Brien, S. J.) 4110–4142 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1995).
34. Peirce, J. L., Lu, L., Gu, J., Silver, L. M. & Williams, R. W. A new set of *Bxd* recombinant inbred lines from advanced intercross populations in mice. *BMC Genet.* **5**, 7 (2004).
35. Williams, R. W. *et al.* Genetic structure of the LXS panel of recombinant inbred mouse strains: a powerful resource for complex trait analysis. *Mamm. Genome* **15**, 637–647 (2004).
36. Darvasi, A. & Soller, M. Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics* **141**, 1199–1207 (1995).
37. Flint, J., Valdar, W., Shifman, S. & Mott, R. Strategies for mapping and cloning quantitative trait genes in rodents. *Nature Rev. Genet.* **6**, 271–286 (2005).
An up-to-date review on methods for mapping quantitative traits in animals models.
38. Valdar, W. *et al.* Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nature Genet.* **38**, 879–887 (2006).
39. Flaherty, L. in *The Mouse in Biomedical Research; History, Genetics, and Wild Mice* (eds Foster, H. L., Small, J. D. & Fox, J. G.) 215–222 (Academic, New York, 1981).
40. Markel, P. *et al.* Theoretical and empirical issues for marker-assisted breeding of congenic mouse strains. *Nature Genet.* **17**, 280–284 (1997).
41. Davis, R. C. *et al.* A genome-wide set of congenic mouse strains derived from DBA/2J on a C57BL/6J background. *Genomics* **86**, 259–270 (2005).
42. Iakoubova, O. A. *et al.* Genome-tagged mice (GTM): two sets of genome-wide congenic strains. *Genomics* **74**, 89–104 (2001).
43. Nadeau, J. H., Singer, J. B., Matin, A. & Lander, E. S. Analysing complex genetic traits with chromosome substitution strains. *Nature Genet.* **24**, 221–225 (2000).
44. Bevova, M. R., Aulchenko, Y. S., Aksu, S., Renne, U. & Brockmann, G. A. Chromosome-wise dissection of the genome of the extremely big mouse line DU6i. *Genetics* **172**, 401–410 (2006).
45. Belknap, J. K. Chromosome substitution strains: some quantitative considerations for genome scans and fine mapping. *Mamm. Genome* **14**, 723–732 (2003).
46. Jaenisch, R. Germ line integration and Mendelian transmission of the exogenous Moloney leukemia virus. *Proc. Natl Acad. Sci. USA* **73**, 1260–1264 (1976).
47. Doetschman, T. *et al.* Targetted correction of a mutant *Hprt* gene in mouse embryonic stem cells. *Nature* **330**, 576–578 (1987).
48. Thomas, K. R. & Capecchi, M. R. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51**, 503–512 (1987).
49. Kwan, K. M. Conditional alleles in mice: practical considerations for tissue-specific knockouts. *Genesis* **32**, 49–62 (2002).
50. Austin, C. P. *et al.* The knockout mouse project. *Nature Genet.* **36**, 921–924 (2004).
51. O'Brien, T. & Woychik, R. Our small relative. *Nature Genet.* **33**, 3–4 (2003).
52. Dipetrillo, K., Wang, X., Stylianou, I. M. & Paigen, B. Bioinformatics toolbox for narrowing rodent quantitative trait loci. *Trends Genet.* **21**, 683–692 (2005).
53. Kanehisa, M. & Bork, P. Bioinformatics in the post-sequence era. *Nature Genet.* **33**, S305–S310 (2003).
54. Baldarelli, R. M. *et al.* Connecting sequence and biology in the laboratory mouse. *Genome Res.* **13**, 1505–1519 (2003).
55. Karp, P. D., Krummenacker, M., Paley, S. & Wagg, J. Integrated pathway–genome databases and their role in drug discovery. *Trends Biotechnol.* **17**, 275–281 (1999).
56. Hubbard, T. *et al.* The Ensembl genome database project. *Nucleic Acids Res.* **30**, 38–41 (2002).
57. Hinrichs, A. S. *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34**, D590–D598 (2006).
58. Karolchik, D. *et al.* The UCSC TABLE Browser data retrieval tool. *Nucleic Acids Res.* **32**, D493–D496 (2004).
59. Bult, C. J. *et al.* The Mouse Genome Database (MGD): integrating biology with the genome. *Nucleic Acids Res.* **32**, D476–D481 (2004).
60. Eppig, J. T. *et al.* The Mouse Genome Database (MGD): from genes to mice — a community resource for mouse biology. *Nucleic Acids Res.* **33**, D471–D475 (2005).
61. Harris, M. A. *et al.* The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32**, D258–D261 (2004).
62. Ashurst, J. L. *et al.* The Vertebrate Genome Annotation (VEGA) database. *Nucleic Acids Res.* **33**, D459–D465 (2005).
63. Bogue, M. Mouse Phenome Project: understanding human biology through mouse genetics and genomics. *J. Appl. Physiol.* **95**, 1335–1337 (2003).
64. Peters, L. L. *et al.* Large scale, high throughput screening for coagulation and hematologic phenotypes in mice. *Physiol. Gen.* **11**, 185–193 (2002).
65. McClurg, P., Pletcher, M. T., Wiltshire, T. & Su, A. I. Comparative analysis of haplotype association mapping algorithms. *BMC Bioinformatics* **7**, 61 (2006).
66. Pletcher, M. T. *et al.* Use of a dense single nucleotide polymorphism map for *in silico* mapping in the mouse. *PLoS Biol.* **2**, 2159–2169 (2004).

67. Gailus-Durner, V. *et al.* Introducing the German Mouse Clinic: open access platform for standardized phenotyping. *Nature Methods* **2**, 403–404 (2005).
68. Schneider, I. *et al.* Systematic, standardized and comprehensive neurological phenotyping of inbred mice strains in the German Mouse Clinic. *J. Neurosci. Methods* **157**, 82–90 (2006).
69. Rowe, L. B. *et al.* Maps from two interspecific backcross DNA panels available as a community genetic mapping resource. *Mamm. Genome* **5**, 253–274 (1994).
70. Moran, J. L. *et al.* Utilization of a whole genome SNP panel for efficient genetic mapping in the mouse. *Genome Res.* **16**, 436–440 (2006).
71. Petkov, P. M. *et al.* An efficient SNP system for mouse genome scanning and elucidating strain relationships. *Genome Res.* **14**, 1806–1811 (2004).
72. Petkov, P. M. *et al.* Development of a SNP genotyping panel for genetic monitoring of the laboratory mouse. *Genomics.* **83**, 902–911 (2004).
73. Pennacchio, L. A. Insights from human/mouse genome comparisons. *Mamm. Genome* **14**, 429–436 (2003).
74. Sugiyama, F. *et al.* Concordance of murine quantitative trait loci for salt-induced hypertension with rat and human loci. *Genomics* **71**, 70–77 (2001).
75. Wang, X., Ishimori, N., Korstanje, R., Rollins, J. & Paigen, B. Identifying novel genes for atherosclerosis through mouse–human comparative genetics. *Am. J. Hum. Genet.* **77**, 1–15 (2005).
76. Klein, R. F. Genetic regulation of bone mineral density in mice. *J. Musculoskelet. Neuronal Interact.* **2**, 232–236 (2002).
77. Korstanje, R. & DiPetrillo, K. Unraveling the genetics of chronic kidney disease using animal models. *Am. J. Physiol. Renal Physiol.* **287**, F347–F352 (2004).
78. Wang, X. & Paigen, B. Genetics of variation in HDL cholesterol in humans and mice. *Circ. Res.* **96**, 27–42 (2005).
79. Wang, X. & Paigen, B. Genome-wide search for new genes controlling plasma lipid concentrations in mice and humans. *Curr. Opin. Lipidol.* **16**, 127–137 (2005).
80. Wang, X. *et al.* Positional identification of *TNFSF4*, encoding OX40 ligand, as a gene that influences atherosclerosis susceptibility. *Nature Genet.* **37**, 365–372 (2005).
81. Hillebrandt, S. *et al.* Complement factor 5 is a quantitative trait gene that modifies liver fibrogenesis in mice and humans. *Nature Genet.* **37**, 835–843 (2005).
82. Ueda, H. *et al.* Association of the T-cell regulatory gene *CTLA4* with susceptibility to autoimmune disease. *Nature* **423**, 506–511 (2003).
83. Korstanje, R. & Paigen, B. From QTL to gene: the harvest begins. *Nature Genet.* **31**, 235–236 (2002).
84. Wade, C. M. *et al.* The mosaic structure of variation in the laboratory mouse genome. *Nature* **420**, 574–578 (2002).
A description of the mosaic structure of the mouse genome.
85. Frazer, K. A. *et al.* Segmental phylogenetic relationships of inbred mouse strains revealed by fine-scale analysis of sequence variation across 4.6 Mb of mouse genome. *Genome Res.* **14**, 1493–1500 (2004).
86. Wiltshire, T. *et al.* Genome-wide single-nucleotide polymorphism analysis defines haplotype patterns in mouse. *Proc. Natl Acad. Sci. USA* **100**, 3380–3385 (2003).
87. Sen, S. & Churchill, G. A. A statistical framework for quantitative trait mapping. *Genetics* **159**, 371–387 (2001).
88. Peters, L. L. *et al.* Identification of quantitative trait loci that modify the severity of hereditary spherocytosis in wan, a new mouse model of band-3 deficiency. *Blood* **103**, 3235–3240 (2004).
89. Ikeda, A. *et al.* Genetic modification of hearing in tubby mice: evidence for the existence of a major gene (*Moth1*) which protects tubby mice from hearing loss. *Hum. Mol. Genet.* **8**, 1761–1767 (1999).
90. Churchill, G. A. & Doerge, R. W. Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963–971 (1994).
This paper introduced permutation analysis into the field of quantitative genetics.
91. Lander, E. S. & Botstein, D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185–199 (1989).
A classic paper that introduced the modern statistical approach to quantitative trait analysis. A challenging but essential reference.
92. Peters, L. L. *et al.* Quantitative trait loci for baseline erythroid traits. *Mamm. Genome* **17**, 298–309 (2006).
93. Peters, L. L. *et al.* Quantitative trait loci for baseline white blood cell count, platelet count, and mean platelet volume. *Mamm. Genome* **16**, 749–763 (2005).
94. Bath, P., Algert, C., Chapman, N. & Neal, B. Association of mean platelet volume with risk of stroke among 3134 individuals with history of cerebrovascular disease. *Stroke* **35**, 622–626 (2004).
95. Castro, O. *et al.* The acute chest syndrome in sickle cell disease: incidence and risk factors. The Cooperative Study of Sickle Cell Disease. *Blood* **84**, 643–649 (1994).
96. de Labry, L. O., Campion, E. W., Glynn, R. J. & Vokonas, P. S. White blood cell count as a predictor of mortality: results over 18 years from the Normative Aging Study. *J. Clin. Epidemiol.* **43**, 153–157 (1990).
97. Martin, J. F., Bath, P. M. & Burr, M. L. Mean platelet volume and myocardial infarction. *Lancet* **339**, 1000–1001 (1992).
98. Martin, J. F., Bath, P. M. & Burr, M. L. Influence of platelet size on outcome after myocardial infarction. *Lancet* **338**, 1409–1411 (1991).
99. Miller, S. T. *et al.* Prediction of adverse outcomes in children with sickle cell disease. *N. Engl. J. Med.* **342**, 83–89 (2000).
100. Platt, O. S. The acute chest syndrome of sickle cell disease. *N. Engl. J. Med.* **342**, 1904–1907 (2000).
101. Platt, O. S. *et al.* Mortality in sickle cell disease. Life expectancy and risk factors for early death. *N. Engl. J. Med.* **330**, 1639–1644 (1994).
102. Platt, O. S. *et al.* Pain in sickle cell disease. Rates and risk factors. *N. Engl. J. Med.* **325**, 11–16 (1991).
103. Darvasi, A. & Soller, M. A simple method to calculate resolving power and confidence interval of QTL map location. *Behav. Genet.* **27**, 125–132 (1997).
104. Zimdahl, H., Kreitler, T., Gosele, C., Ganten, D. & Hubner, N. Conserved synteny in rat and mouse for a blood pressure QTL on human chromosome 17. *Hypertension* **39**, 1050–1052 (2002).
105. Vitt, U. *et al.* Identification of candidate disease genes by EST alignments, synteny, and expression and verification of Ensembl genes on rat chromosome 1q43–54. *Genome Res.* **14**, 640–650 (2004).
106. Parsons, C. A. *et al.* Interspecies synteny mapping identifies a quantitative trait locus for bone mineral density on human chromosome Xp22. *Hum. Mol. Genet.* **14**, 3141–3148 (2005).
107. Li, R., Lyons, M. A., Wittenburg, H., Paigen, B. & Churchill, G. A. Combining data from multiple inbred line crosses improves the power and resolution of quantitative trait loci mapping. *Genetics* **169**, 1699–1709 (2005).
108. DiPetrillo, K. *et al.* Genetic analysis of blood pressure in C5HiHeJ and SWRJ mice. *Physiol. Genomics* **17**, 215–220 (2004).
109. Ishimori, N. *et al.* Quantitative trait loci that determine BMD in C57BL/6J and 129S1/SvImJ inbred mice. *J. Bone Miner. Res.* **21**, 105–112 (2006).
110. Wang, X., Korstanje, R., Higgins, D. & Paigen, B. Haplotype analysis in multiple crosses to identify a QTL gene. *Genome Res.* **14**, 1767–1772 (2004).
111. de Botton, S. *et al.* Platelet formation is the consequence of caspase activation within megakaryocytes. *Blood* **100**, 1310–1317 (2002).
112. Hartwig, J. & Italiano, J. Jr. The birth of the platelet. *J. Thromb. Haemost.* **1**, 1580–1586 (2003).
113. Italiano, J. E. & Shivdasani, R. A. Megakaryocytes and beyond: the birth of platelets. *J. Thromb. Haemost.* **1**, 1174–1182 (2003).
114. Lecine, P. *et al.* Mice lacking transcription factor NF-E2 provide *in vivo* validation of the proplatelet model of thrombocytopoiesis and show a platelet production defect that is intrinsic to megakaryocytes. *Blood* **92**, 1608–1616 (1998).
115. Lecine, P., Italiano, J. E., Kim, S. W., Villeval, J. L. & Shivdasani, R. A. Hematopoietic-specific β1 tubulin participates in a pathway of platelet biogenesis dependent on the transcription factor NF-E2. *Blood* **96**, 1366–1373 (2000).
116. Tiwari, S. *et al.* A role for Rab27b in NF-E2-dependent pathways of platelet formation. *Blood* **102**, 3970–3979 (2003).
117. Levin, J. & Bessman, J. D. The inverse relation between platelet volume and platelet number. Abnormalities in hematologic disease and evidence that platelet size does not correlate with platelet age. *J. Lab. Clin. Med.* **101**, 295–307 (1983).
118. Rosa, G. J., de Leon, N. & Rosa, A. A review of microarray experimental design strategies for genetical genomics studies. *Physiol. Genomics* 19 Sep 2006 [epub ahead of print].
119. Lum, P. Y. *et al.* Elucidating the murine brain transcriptional network in a segregating mouse population to identify core functional modules for obesity and diabetes. *J. Neurochem.* **97**, 550–562 (2006).
120. Doss, S., Schadt, E. E., Drake, T. A. & Lusis, A. J. *Cis*-acting expression quantitative trait loci in mice. *Genome Res.* **15**, 681–691 (2005).
121. Drake, T. A., Schadt, E. E. & Lusis, A. J. Integrating genetic and gene expression data: application to cardiovascular and metabolic traits in mice. *Mamm. Genome* **17**, 466–479 (2006).
122. Yaguchi, H., Togawa, K., Moritani, M. & Itakura, M. Identification of candidate genes in the type 2 diabetes modifier locus using expression QTL. *Genomics* **85**, 591–599 (2005).
123. Valdar, W., Flint, J. & Mott, R. Simulating the Collaborative Cross: power of quantitative trait loci detection and mapping resolution in large sets of recombinant inbred strains of mice. *Genetics* **172**, 1783–1797 (2006).
124. Churchill, G. A. *et al.* The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nature Genet.* **36**, 1133–1137 (2004).
This paper describes the design and objective of the Collaborative Cross.
125. Grupe, A. *et al. In silico* mapping of complex disease-related traits in mice. *Science* **292**, 1915–1918 (2001).
The first description of *in silico* mapping of complex traits
126. Smith, J. D. *et al. In silico* quantitative trait locus map for atherosclerosis susceptibility in apolipoprotein E-deficient mice. *Arterioscler. Thromb. Vasc. Biol.* **23**, 117–122 (2003).
127. Li, R. *et al.* Structural model analysis of multiple quantitative traits. *PLoS Genet.* **2**, e114 (2006).
A breakthrough paper describing the use of structural modelling to identify cause-and-effect relationships among genes and phenotypes.
128. Wang, X. *et al.* Using advanced intercross lines for high-resolution mapping of HDL cholesterol quantitative trait loci. *Genome Res.* **13**, 1654–1664 (2003).

Acknowledgements
Thanks to the US National Institute of General Medical Sciences and the US National Heart, Lung and Blood Institute.

Competing interests statement
The authors declare no competing financial interests.

DATABASES
The following terms in this article are linked online to:
Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
ANK1 | Ank2 | Apc | db | Hc | Mom1 | Nfe2 | ob | Stesp1

FURTHER INFORMATION
Deltagen: http://www.deltagen.com
Deltagen and Lexicon Knockout Mice and Phenotypic Data: http://www.nih.gov/science/models/mouse/deltagenlexicon/index.html
Eumorphia EMPReSS: http://empress.har.mrc.ac.uk/EMPReSS/servlet/EMPReSS.Frameset
European Conditional Mouse Mutagenesis Program: http://www.eucomm.org
Genome Variation Server: http://pga.mbt.washington.edu/
Genomics Institute of the Novartis Research Foundation: http://web.gnf.org
German Mouse Clinic: http://www.mouseclinic.de
International HapMap Project: http://www.hapmap.org
Knockout Mouse Project: http://www.nih.gov/science/models/mouse/knockout/index.html
Lee Silver's Mouse Genetics: http://www.informatics.jax.org/silver
Lexicon Genetics: http://www.lexicon-genetics.com/index.php
North American Conditional Mouse Mutagenesis Project: http://www.genomecanada.ca
Perlegen Sciences: http://www.perlegen.com
PhyaGen: http://pga.mcw.edu
QTL Archive: http://www.jax.org/staff/churchill/labsite/datasets/qtlarchive/index.html
Roche: http://www.roche.com/home.html
Sanger Institute: http://www.sanger.ac.uk
The Institute for Genome Research: http://www.tigr.org
The Jackson Laboratory: http://jaxmice.jax.org/index.html
Trans-NIH Mouse Initiatives: http://www.nih.gov/science/models/mouse
Access to this links box is available online.